

# Feature Fusion Reveals Slow and Fast Visual Memories

Frank Scharnowski<sup>1</sup>, Frouke Hermens<sup>1</sup>, Thomas Kammer<sup>2</sup>,  
Haluk Ögmen<sup>3</sup>, and Michael H. Herzog<sup>1</sup>

## Abstract

■ Although the visual system can achieve a coarse classification of its inputs in a relatively short time, the synthesis of qualia-rich and detailed percepts can take substantially more time. If these prolonged computations were to take place in a retinotopic space, moving objects would generate extensive smear. However, under normal viewing conditions, moving objects appear relatively sharp and clear, suggesting that a substantial part of visual short-term memory takes place at a *nonretinotopic* locus. By using a retinotopic feature fusion and

a nonretinotopic feature attribution paradigm, we provide evidence for a relatively fast retinotopic buffer and a substantially slower nonretinotopic memory. We present a simple model that can account for the dynamics of these complementary memory processes. Taken together, our results indicate that the visual system can accomplish temporal integration of information while avoiding smear by breaking off sensory memory into fast and slow components that are implemented in retinotopic and nonretinotopic loci, respectively. ■

## INTRODUCTION

“The hour strikes on a clock ... and yet I only have to turn my attention backwards, to count up the four strokes which have already sounded, and add them to those which I hear.” Henri Bergson (1889) was one of the first to distinguish between the extrinsic time course of events and their mental counterparts (see also Metzger, 1932; Rubin, 1930; Fröhlich, 1929; Cattell, 1885). Since then, philosophers, psychologists, and neurobiologists have addressed the various aspects of intrinsic timing, such as the time for a stimulus to reach awareness and perceptual duration (e.g., Bachmann, 2000). Several aspects of the temporal dynamics of visual processes were studied, including temporal (Francis, 1996; Di Lollo, 1977; Pollack, 1973; Eriksen & Collins, 1967) and trans-saccadic integration (Melcher, 2005; Melcher & Morrone, 2003; Irwin, Brown, & Sun, 1988; McRae, Butler, & Poppel, 1987; Bridgeman & Mayer, 1983; Davidson, Fox, & Dick, 1973), the duration of a conscious percept (Efron, 1967, 1970a, 1970b), and the time course of neural and perceptual processing of a stimulus (Macknik & Martinez-Conde, 2004; Lamme, Zipser, & Spekreijse, 2002; Super, Spekreijse, & Lamme, 2001a, 2001b; VanRullen & Thorpe, 2001b; Macknik, Martinez-Conde, & Haglund, 2000; Macknik & Livingstone, 1998; Thorpe, Fize, & Marlot, 1996; Bullier & Nowak, 1995). The investigations led to the conceptualization of different visual memories, such as “visible persistence” and “iconic memory” (Haber, 1983; Coltheart, 1980; Neisser, 1967; Sperling, 1960).

From a system-theoretical point of view, the memory of a system can be not only an essential part of its computational structure, but can also be a side effect of its physical limitations. For example, all physical systems have inertia that makes them unable to respond instantaneously to changes in their input. In other words, when the input changes, a “memory” of the previous state continues to persist. Our inability to perceive flicker beyond a certain critical frequency is a direct consequence of this inertia. On the other hand, a slow response time helps the system to integrate its input, a characteristic that can be beneficial in increasing the sensitivity of the system under low-light conditions.

Although the visual system can achieve a coarse classification of its inputs in a relatively short time (Thorpe, Delorme, & van Rullen, 2001; van Rullen & Thorpe, 2001a, 2001b; Thorpe et al., 1996), the synthesis of qualia-rich and detailed percepts can take considerably longer (Brincat & Connor, 2006; Liu, Harris, & Kanwisher, 2002; Sugase, Yamane, Ueno, & Kawano, 1999). If this prolonged computation were to take place in a retinotopic space, moving objects would generate extensive smear. However, under normal viewing conditions, moving objects typically appear relatively sharp and clear (Tong, Patel, & Bedell, 2005; Chen, Ögmen, & Bedell, 1995; Westerink & Teunissen, 1995; Castet, 1994; Farrell, Pavel, & Sperling, 1990; Morgan & Benton, 1989; Hogben & Di Lollo, 1985; Burr, 1980; Dixon & Hammond, 1972), even if the moving object is physically blurred to some extent (Hammett, Georgeson, & Gorea, 1997; Bex, Edgar, & Smith, 1995; Ramachandran, Rao, & Vidyasagar, 1974).

One way to deal with this “processing time versus smear” trade-off is to have a relatively fast retinotopic

<sup>1</sup>École Polytechnique Fédérale de Lausanne (EPFL), Switzerland,

<sup>2</sup>University of Ulm, Germany, <sup>3</sup>University of Houston

buffer that would provide the necessary time for neurons in the early visual cortex to register incoming stimuli, and to incorporate the substantial part of visual memory at a *nonretinotopic* locus where object-based information can be integrated and processed. Here, we examine this possibility by combining two psychophysical paradigms.

First, by analyzing temporal-order effects in a feature fusion paradigm we provide evidence for a relatively fast retinotopic buffer that allows integration of spatially local information. Second, using a Ternus–Pikler paradigm, we provide evidence for a slow nonretinotopic buffer that allows integration of information from different retinotopic loci. Finally, we speculate about the neural processes linking these two complementary mechanisms.

## EXPERIMENT 1: TEMPORAL DYNAMICS OF RETINOTOPIC FEATURE FUSION

In Experiment 1, we introduce an experimental paradigm to obtain an estimate of retinotopic integration time. The basic idea is to present elements with conflicting feature information at the same retinotopic locus. When presented very briefly and in rapid succession, the elements are perceptually fused and perceived as one coherent object. As elements to be fused, we presented verniers with opposite offset directions. A vertical line vernier is a pair of vertical bars that are spatially offset in the horizontal direction (Figure 1). The first vernier is immediately followed by a second vernier of the same duration and the same offset size but opposite offset direction. Because of its opposite offset direction,

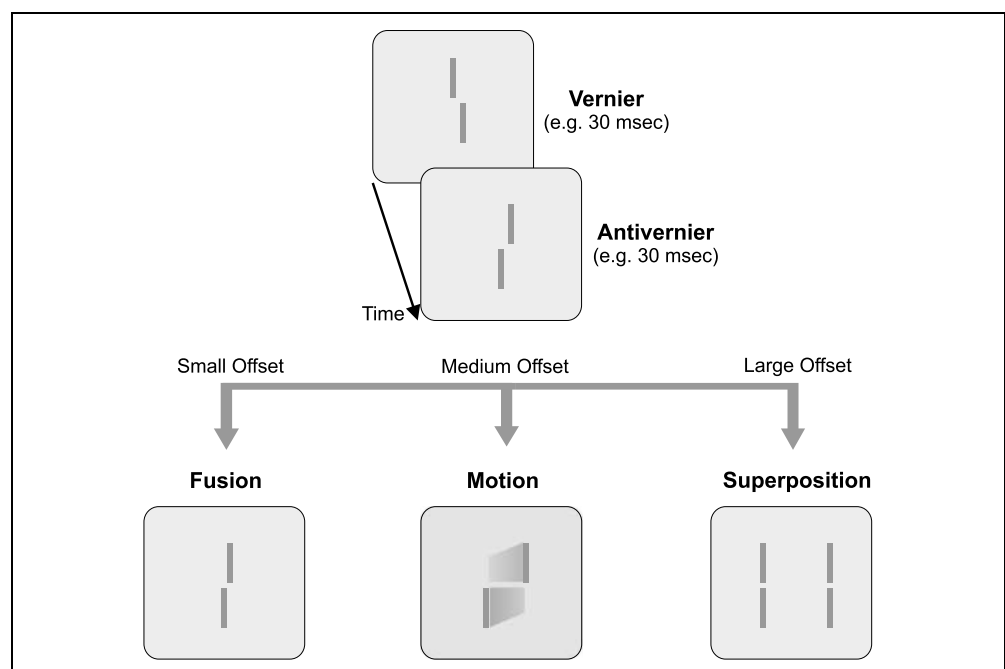
we refer to the second vernier as “antivernier” (to the first vernier). Hence, if the vernier is offset to the left, the antivernier is offset to the right and vice versa. Depending on the duration and the offset size of the verniers, three possible percepts can occur: feature fusion, motion, and spatial superposition (Figure 1).

If vernier and antivernier are presented for short durations and have offset sizes smaller than  $\sim 1'$  (strongly depending on the observer), feature fusion occurs: Observers experience only one fused vernier with one offset direction and are not aware of the two individual verniers. When the offsets are in the range between  $\sim 1'$  and  $\sim 2'$ , participants report seeing motion, that is, a vernier whose offset direction appears to flip from the vernier to the antivernier direction. Increasing the offset of the elements further ( $> 2'$ ) causes superposition of the vernier elements: Four widely spaced lines appear to be displayed simultaneously. Because we are interested in feature fusion, offset sizes in the experiment were adjusted individually to avoid motion and superposition.

The offset of the fused vernier is a result of integrating the offsets of vernier and antivernier. By changing the respective durations, luminances, and spatial offsets of vernier and antivernier, the perceived offset of the fused vernier can be manipulated.

To determine the temporal dynamics of feature fusion, we presented different sequences of verniers and antiverniers of equal overall physical energy: In all conditions, the product of luminance and duration of all verniers combined is the same as for all antiverniers combined. Because we varied only the sequential order of the elements, any changes in performance reflect temporal effects and can be used to directly probe the dynamics of the retinotopic buffer.

**Figure 1.** A vernier is immediately followed by its antivernier. Depending on the duration and the offset size of the verniers, three different percepts are possible. For small offset sizes and short durations, the two verniers are *fused*; that is, only one vernier is perceived. Medium offset sizes result in *motion* percepts. For larger offset sizes, verniers are perceived simultaneously and spatially *superimposed*.



## Methods

### Participants

Four observers (2 women; aged 21–29 years; normal or corrected-to-normal visual acuity) gave informed written consent for participation in the study, which was approved by the local ethics committee. All observers but one were naive to the purpose of the study. They were paid 20 CHF/hr (approximately €13/hr) for their participation.

### Stimuli

White stimuli were displayed in the center of an  $x$ - $y$  display (Tektronix 608, equipped with a P11 phosphor, controlled by a PC via fast 16-bit DA converters; dot pitch of 200  $\mu\text{m}$  at a dot rate of 1 MHz; 200 Hz refresh rate) on a dark background at 80  $\text{cd}/\text{m}^2$ . Participants observed the stimuli from a distance of 2 m in a room dimly illuminated by a background light ( $\sim 0.5$  lx). Verniers were composed of two vertical bars that were slightly displaced in the horizontal direction. Vernier segments were 10' long and separated by a vertical gap of 1'. Hence, the total height of the vernier was 21'. The direction of the vernier offset (left or right) was chosen randomly for each trial. The antiverniers had the same spatial parameters as the verniers except for opposite offset direction. If, for example, a vernier was offset to the left, antiverniers were offset to the right and vice versa. The spatial offsets were adjusted individually and ranged from 40" to 50".

### Procedure

Verniers were offset to the left or right randomly (antiverniers to the right or left accordingly). In a binary task, we asked observers to indicate the perceived offset of the *fused* vernier. We presented either a sequence of a vernier and an antivernier or of a vernier, an antivernier, and a vernier. Each sequence was presented in blocks of 80 trials of which the order was randomized across participants. The sequences were repeated in opposite order, yielding a total of 160 trials for every condition. The *overall* duration of the verniers and the antivernier was identical in all sequences. Hence, their physical energy was identical throughout all conditions.

### Analysis

Performance was quantified as the percentage of responses in which the perceived offset of the fused vernier corresponds to the direction of the offset of the vernier presented first. Values above 50% indicate that the vernier dominates performance; values below 50% indicate that the antivernier dominates.

## Results and Discussion

### Temporal-order Effects in Retinotopic Feature Fusion

Four sequences of verniers and antiverniers were presented while keeping their total physical energy constant; that is, luminance and total duration of all verniers and those of the antivernier were equal. Subjects perceived the sequence of verniers and antiverniers as only one fused vernier, whose offset direction they indicated. In the first condition, a single vernier was followed by a single antivernier. In this condition, the antivernier clearly dominates ( $V_{30 \text{ msec}}\text{-}AV_{30 \text{ msec}}$ ; Herzog, Parish, Koch, & Fahle, 2003). In the second condition, we subdivided the vernier into two parts, of which one was presented before and the other after the antivernier. When only a small share of the verniers (5 msec) is presented after the antivernier, the antivernier still dominates ( $V_{25 \text{ msec}}\text{-}AV_{30 \text{ msec}}\text{-}V_5 \text{ msec}$ ). The third condition shows that the offset dominance of the fused vernier is almost balanced, when the vernier is split into two equal parts of 15 msec each ( $V_{15 \text{ msec}}\text{-}AV_{30 \text{ msec}}\text{-}V_{15 \text{ msec}}$ ). The fourth condition shows that the verniers dominate when increasing the share of the vernier element after the antivernier to 25 msec ( $V_5 \text{ msec}\text{-}AV_{30 \text{ msec}}\text{-}V_{25 \text{ msec}}$ ).

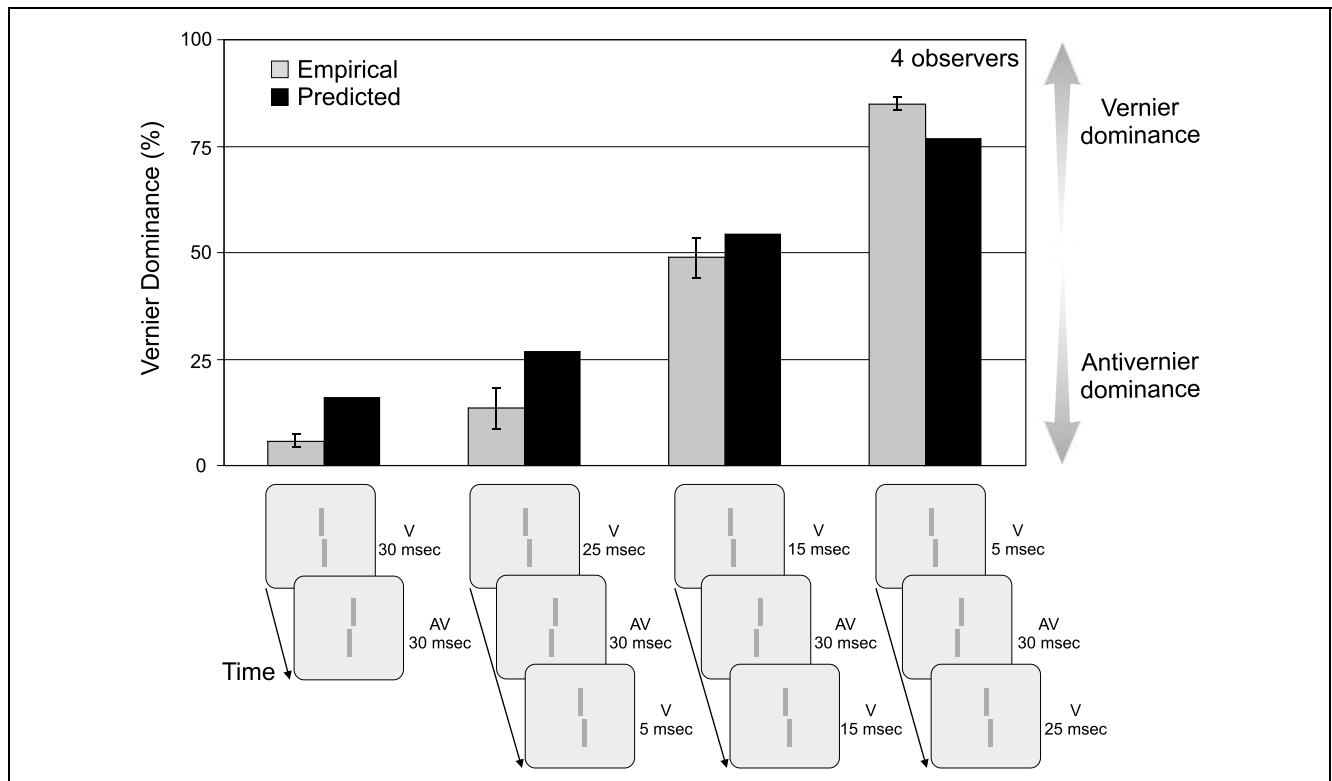
Our results indicate that any level of performance can be achieved by arranging the sequence of verniers and antiverniers appropriately—even though the total physical energy of verniers and antiverniers is identical. This suggests that later elements of the sequence contribute more to the perceived offset of the fused vernier than earlier ones. Hence, the temporal order of events is critical in determining how features fuse.

### A Simple Dynamic Model for Fast Retinotopic Feature Fusion

The temporal-order effects in feature fusion cannot be captured by energy-based models that simply summate the energy of the individual elements over time. Such models would predict 50% vernier dominance for all four conditions. Therefore, a model is required that dynamically links the activity of neurons that code the offset information of the individual verniers. Surprisingly, a simple so-called passive resistor-capacitor (RC) model for neural dynamics can explain the results of Figure 2. A schematic of this model is shown in Figure 3.

Two neurons (labeled A and B) collect information for either a left or a right offset of the vernier. A third neuron (C) combines the information from neurons A and B. An RC circuit acts as a leaky integrator allowing each neuron to integrate its inputs at a rate that is identical to its decay rate. For simplicity, we assumed that the activation in all three neurons decays according to the same decay rate constant  $\tau$ . This decay is represented in Figure 3 as recurrent negative feedback loops.

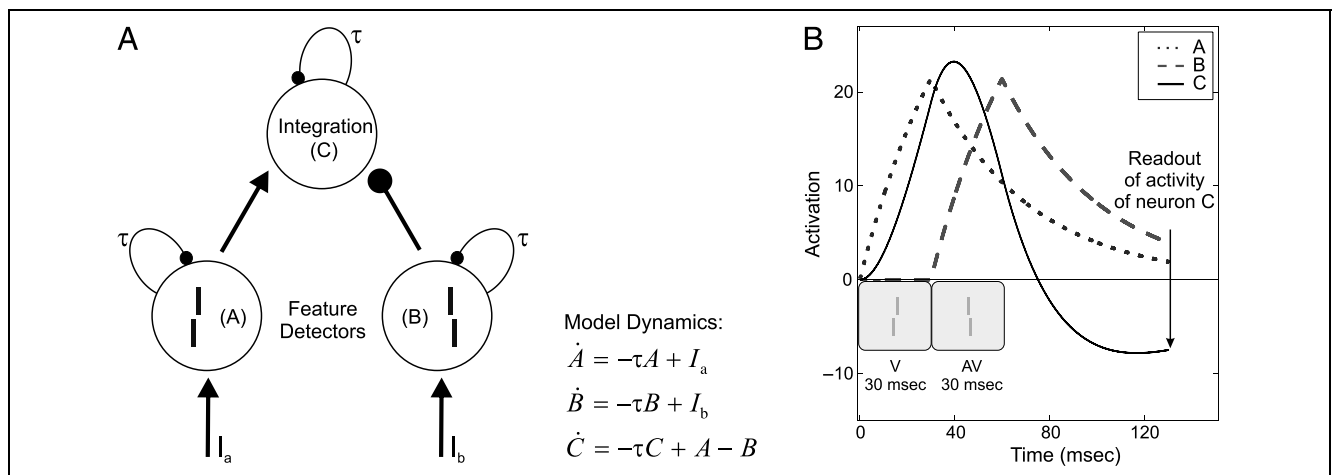
The three free parameters of the model (decay rate constant  $\tau$ , readout time, and the slope of the linking



**Figure 2.** Vernier dominance of four sequences of verniers and antiverniers with constant overall physical energy (light gray). The later the elements are presented, the higher seems to be their impact on performance. In black, predictions of an RC model (Figure 3) are shown, which will be introduced in the section “A simple dynamic model for fast retinotopic feature fusion.” Error bars indicate *SEM*. V = vernier; AV = antivernier.

sigmoid function  $\sigma$ ) were fit using the data of Experiment 2 and another 30 different sequences of several verniers and antiverniers with individual vernier presentation times ranging from 5 to 60 msec (unpublished

data). The model predictions for all 34 conditions agree quite well with the behavioral results ( $r^2 = .89$ ). Predictions for the four sequences of verniers and antiverniers are shown in Figure 2 (black bars).



**Figure 3.** Dynamic three-neuron model. (A) The feature detector neurons (A and B) collect evidence for either a left (A) or a right (B) vernier offset direction. Their activation is fed into an integration neuron C. Activity is assumed to decay in all neurons according to the decay rate constant  $\tau$  (negative feedback loops). Positive input is denoted by an arrowhead; negative input as a filled dot at the end of the connecting line. For the duration of a left offset, the input  $I_a$  is set to one. For a right offset  $I_b$  is set to one. If no stimulus is presented both inputs are zero. (B) The computation of neuron C is such that if the evidence for a left offset (A) dominates, the activation of neuron C is positive at the readout time of 70.8 msec after the termination of the last stimulus element. If evidence for a right (B) offset dominates, the activation of neuron C is negative. To map the activation of C to performance, we used a sigmoidal function as a linking hypothesis (see the Appendix for details). V = vernier; AV = antivernier.

By setting the decay in the model to zero, nonleaky energy summation is obtained. This corresponds to a pure integrator model commonly used in spatial vision. Such a model predicts that performance is identical in all conditions and therefore fails to explain why later elements of the sequence are more important than earlier ones.

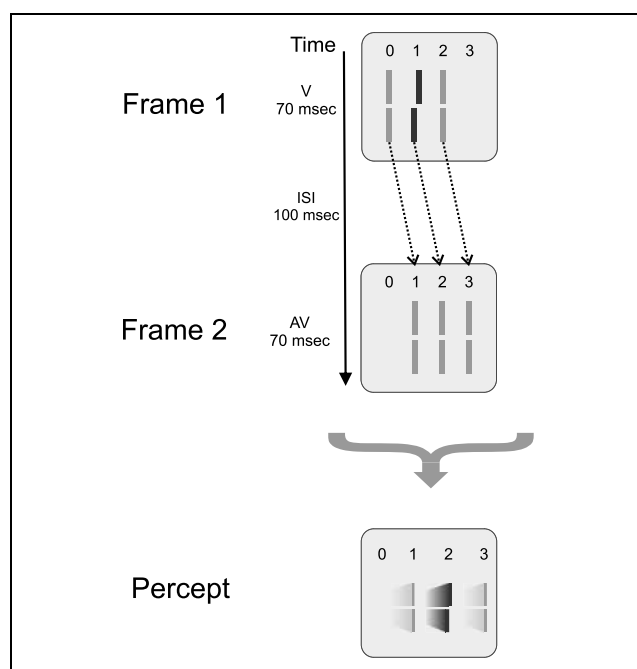
Experiment 1 demonstrated that in feature fusion, elements presented later influence perception more strongly than earlier ones. A possible explanation of this finding is that neural activity related to the spatial vernier offset decays. Possibly, this decay also occurs when only a single vernier is presented and no other elements follow. Despite the decay of this vernier signal, observers are nevertheless able to report the spatial offset of a single vernier. When presenting only one stimulus, this decay cannot be estimated because decay rate and readout time are two independent temporal variables. Whatever the exact mechanisms are, presenting sequences of verniers and antiverniers allows an estimation of the temporal dynamics of their interactions. Hence, our paradigm, like other temporal integration approaches to memory (Di Lollo, 1977; Pollack, 1973; Eriksen & Collins, 1967) reveals otherwise hidden temporal mechanisms. The decay rate constant  $\tau$ , which we have estimated by fitting our model to data of the 34 conditions, was  $0.0291 \text{ msec}^{-1}$ , resulting in a time constant of  $1/0.0291 = 34.36 \text{ msec}$ .

## EXPERIMENT 2: TEMPORAL DYNAMICS OF NONRETINOTOPIC FEATURE ATTRIBUTION

As mentioned in the Introduction, without a fast retinotopic buffer, moving objects would appear highly smeared (Chen et al., 1995). Therefore, curtailing the duration of the retinotopic buffer makes a *nonretinotopic memory* necessary, because processing of complex stimuli may require substantially longer times than that allowed by the fast retinotopic buffer.

To probe the dynamics of nonretinotopic memory, we applied a version of a stimulus known as the “Ternus–Pikler display” (Ternus, 1926; Pikler, 1917). In this display, three elements are presented in a first frame, followed by a blank screen for a certain interstimulus interval (ISI), followed by a second frame with the three elements shifted to the right (Figure 4). For an ISI of 100 msec, a motion percept is elicited in which the three elements of the first frame move *as a group* to match the corresponding three elements of the second frame.

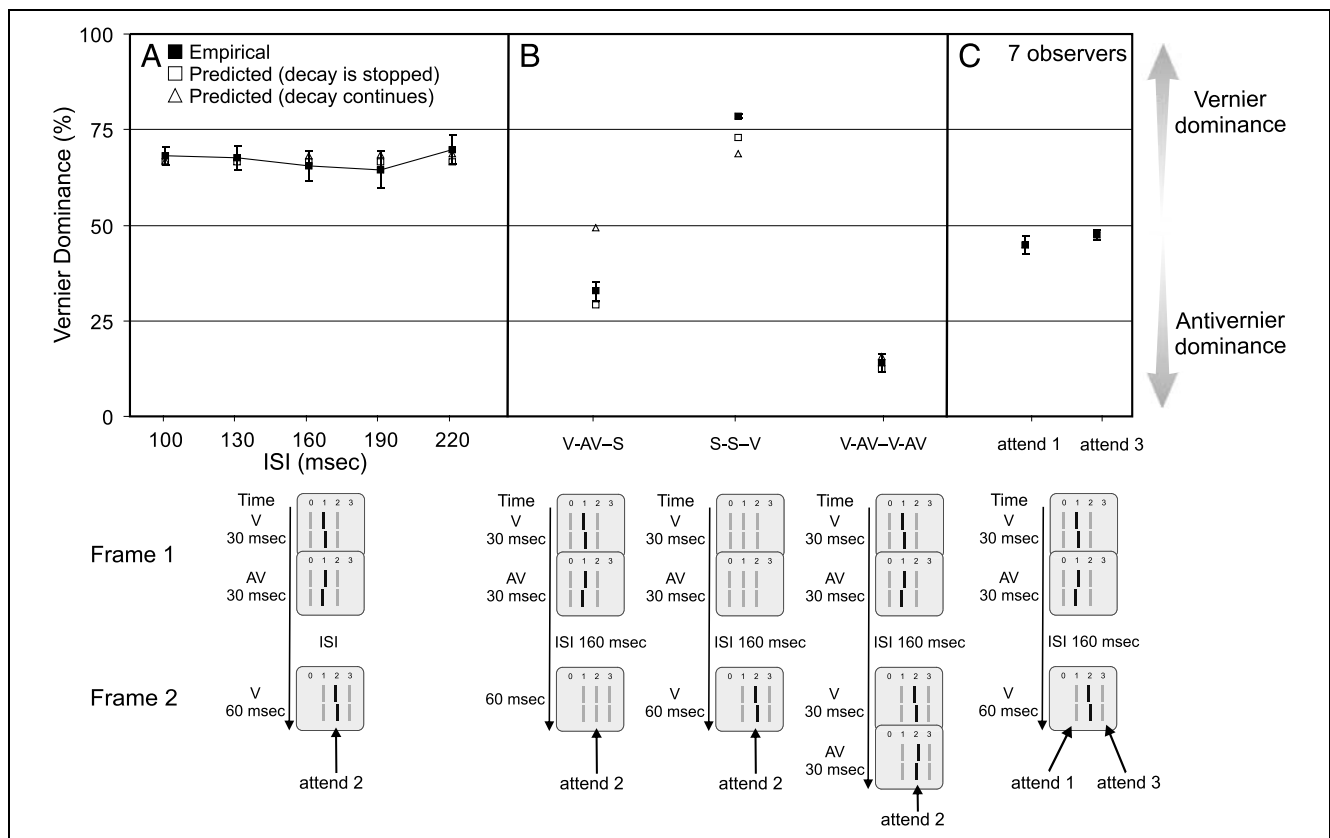
If a spatial offset is added to the central vernier element of the first frame (Figure 4A; Position 1 highlighted in black), this vernier offset is surprisingly perceived at the central element of the second frame, even though at this retinotopic position no offset is presented neither in the first nor in the second frame (Figure 4A; Position 2). This implies that feature attribution is *nonretinotopic* and follows the correspondence of elements



**Figure 4.** Ternus–Pikler display. Three lines are presented in the first frame. When after an ISI of 100 msec these elements are shifted to the right, observers perceive three lines moving as a group. Feature attribution takes place among elements for which a motion correspondence is established (as indicated by the three arrows). For example, a spatial offset of the central element of the first frame (Position 1 highlighted in black) is perceived at the middle element of the second frame (Position 2) even though at this retinotopic position no vernier offset was presented in either of the two frames. V = vernier; AV = antivernier; ISI = interstimulus interval. Luminance of all lines was identical; the black elements are to highlight the position of the offset.

established by motion-induced perceptual grouping (Ögmen, Otto, & Herzog, 2006). This paradigm offers the advantage of studying nonretinotopic feature processing without using eye movement paradigms (e.g., Irwin et al., 1988; McRae et al., 1987; Davidson et al., 1973) and thus eliminates potential interferences from eye-movement-related effects.

We made use of this effect and combined feature fusion at one spatial position with the Ternus–Pikler display. Instead of only presenting one vernier in the first frame (Figure 4; Position 1 highlighted in black), we presented a vernier followed by an antivernier (Figure 5A, first frame; Position 1 highlighted in black). From Experiment 1, we know that the vernier and the antivernier within this first frame are fused. In the second experiment, we show that the fused vernier offsets of the first frame are attributed nonretinotopically over space to integrate with the corresponding vernier of the second frame (Figure 5A, second frame; Position 2 highlighted in black). To integrate the information of the two frames, the fused feature information of the first frame has to be stored until it becomes integrated with the corresponding element of the second frame.



**Figure 5.** Feature fusion and integration over space and time with the Ternus-Pikler display. In the first frame, feature fusion is taking place at the central position. With the onset of the second frame this fused feature information is integrated with the central vernier of the second frame. (A) Observers were asked to report the perceived spatial offset of the attended element of the second frame. (A) Varying the ISI between first and second frame from 100 to 220 msec has no obvious effect on performance, indicating that the fused feature information of the first frame is memorized without significant decay. (B) Here, we show that the performance in (A) is not only based on the offset information of the second frame, but is the result of integrating offset information of both frames. When replacing the central vernier of the second frame with a straight element (V-AV-S), the antivernier dominates. This shows that the fused feature information of the first frame is indeed integrated with the corresponding element of the second frame. Replacing the spatially offset vernier and antivernier of the first (S-S-V) frame with straight elements (V-AV-V-AV) renders the vernier even more dominant than in (A). When presenting a vernier followed by an antivernier in both frames, the antivernier dominates in the first as well as in the second one and, accordingly, the dominance of the antivernier is more pronounced (V-AV-V-AV). (C) When observers attend to the leftmost or the rightmost element of the second frame, they perform close to chance level. This shows that feature integration does not take place according to retinotopic coordinates but among elements whose correspondence is consistent with motion-induced perceptual grouping. Open symbols show predictions of the model simulations, which are described in more detail in the main text. Error bars indicate *SEM*; in some cases they are smaller than symbol size. V = vernier; AV = antivernier; S = straight vernier; ISI = interstimulus interval.

## Methods

### Participants

Seven new observers were tested.

### Stimuli

The stimuli are illustrated in Figure 5. Three main conditions, labeled A, B, and C, were used. The results obtained in each condition are shown in the corresponding panel of Figure 5 along with a depiction of the corresponding stimulus configuration at the bottom of each panel.

In Condition A, the offsets of the vernier and the antivernier in the first frame were set to 50". The horizontal distance between the vernier/antivernier and the two flanking straight verniers was 800". After a variable ISI

(100–220 msec in steps of 30 msec), the second frame was displayed with three elements shifted to the right by 800". The offset of the vernier in the second frame was determined individually so that participants performed at around 65% at an ISI of 100 msec (offset range, 15"–20").

In Conditions B and C the ISI was 160 msec. Condition B contained three different stimulus configurations. In the first configuration, the vernier of the second frame was replaced by a straight vernier (V-AV-S). In the second configuration, the vernier and antivernier of the first frame were both replaced by straight verniers (S-S-V). In the last configuration, the second frame was identical to the first one (V-AV-V-AV).

Condition C was the same as in A with the exception of the ISI, which was kept constant at 160 msec, and the instructions that required the observers to attend

to the leftmost or to the rightmost element of the second frame.

### *Procedure and Design*

Observers were instructed to attend to the central element in the second frame in Conditions A and B, and to the first or third element of the second interval in Condition C. They were asked to report the perceived direction of the vernier offset (left or right) for this attended element by pressing one of two push buttons. Naive observers had no knowledge about where the spatially offset elements were presented and that their responses may have been based on an element that did not contain a spatial offset.

### *Analysis*

Performance was analyzed as in Experiment 1.

## **Results and Discussion**

### *Nonretinotopic Feature Integration*

In the first condition (Figure 5A), we presented a vernier and antivernier at Position 1 in the first frame and a vernier at Position 2 in the second frame. Based on the assumption that retinotopic offset information of the first frame rapidly decays, the vernier element of the second frame should become more dominant with increasing ISI. However, performance is roughly constant across ISIs (Figure 5A). Even for an ISI of 220 msec, vernier dominance remains at the level we aimed at during the normalization procedure with an ISI of 100 msec, that is, around 65%. This indicates that feature information of the first frame is not rapidly decaying but temporarily stored, at least within the range of ISIs used in our experiment.

Performance in the conditions in A would also be constant if the offset information of the first frame has completely vanished and only the vernier of the second frame has been taken into account. Figure 5B shows, however, that the offset information of the first frame has a strong impact on performance. Hence, both frames are integrated. When the vernier of the second frame is replaced by a straight element, the antivernier dominates performance (Figure 5B; V-AV-S). This indicates that the information from the first frame is still available at the moment it becomes integrated with the second frame. To determine the contribution of the second frame, we replaced vernier and antivernier of the first frame by straight elements. In this condition, the vernier of the second frame strongly dominates (Figure 5B; S-S-V). Vernier dominance is more pronounced than in A (Figure 5A; V-AV-V), because no fused feature information with antivernier dominance is present in the first frame. When presenting a vernier followed by an antivernier in both frames (Figure 5B; V-AV-V-AV), anti-

vernier dominance is even stronger than in A, because the antivernier dominates in both frames (Figure 5B; V-AV-V-AV). This illustrates that feature fusion takes place at one spatial location in the first frame and the fused information becomes then integrated over space with the second frame.

To verify that features are integrated over space according to object grouping, we asked observers to attend to the first or third element of the second frame (Figure 5C). Even though the vernier and antivernier of the first frame are presented at the same retinotopic position as the first element of the second frame, observers' responses do not reflect this offset information except for a small leakage of the first element of the second frame that spatially overlaps the vernier and antivernier of the first frame. This leakage is also present in Ögmen et al. (2006) (Figure 5C). Hence, feature integration takes place according to the correspondence of elements established by motion-induced perceptual grouping.

### *Nondecaying Memory as an Explanation for the ISI-invariant Performance in Nonretinotopic Feature Integration*

According to the RC model, information from the first frame has almost completely decayed before the onset of the second frame. This fails to explain antivernier dominance in the condition V-AV-S (Figure 5B; open triangles). Here, the model predicts performance to be at 50%, because feature information of the first frame has already decayed at the moment the second frame is presented, and in the second frame no spatial offset is presented. To account for the results, we propose that information is read out before a complete decay has occurred. To model this, we froze the activations of all three neurons shortly after the termination of the first frame for the duration of the ISI (Figure 5; open rectangles). After the onset of the second frame, the decay continues. With these modifications, the model predictions agree with the experimental results.

## **GENERAL DISCUSSION**

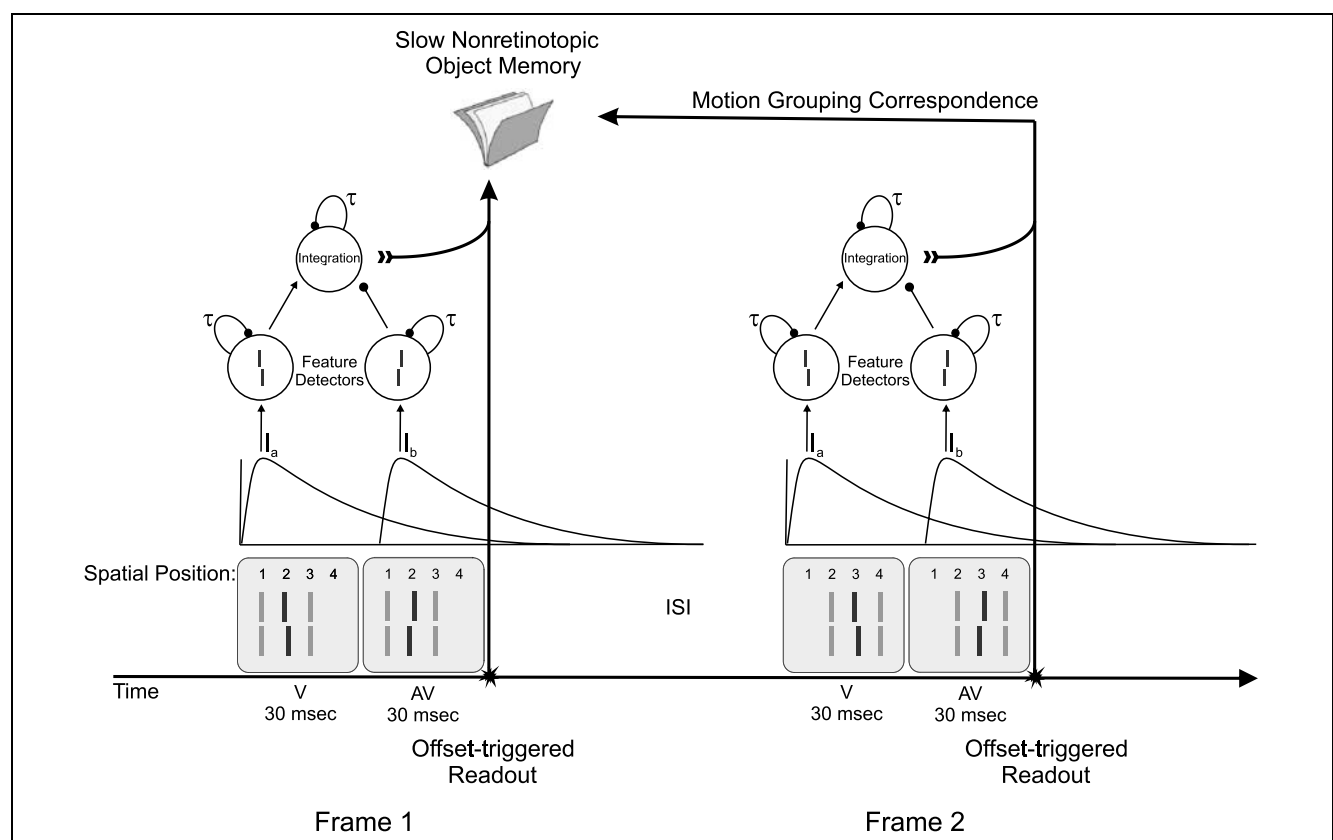
When verniers are presented in rapid succession, they are perceptually fused and their offsets are integrated. Elements that are presented later dominate feature integration. Our findings show that vision is a dynamic process that cannot be captured by static integration models. Such models summate evidence for the vernier and the antivernier, which should cancel each other because their energy was identical (Figure 2). Also, feed-forward models of ultrarapid object categorization (van Rullen & Thorpe, 2001a; Thorpe et al., 1996, 2001) predict dominance of the elements presented first, which is also incompatible with our results.

Feature fusion is not restricted to verniers, but has also been reported in color vision (Yund, Morgan, & Efron, 1983; Efron, 1973), in the auditory (Efron, 1973), and in the tactile domain (Efron, 1973). As with verniers, the trailing elements dominate.

We suggest that the dominance of the later elements may be explained with a rapid decay of neural activity coding these features. The proposed decay rate in our experiments is quite fast ( $\tau = 0.0291 \text{ msec}^{-1}$ , corresponding to a time constant of 34 msec and a half-life time of decay of 24 msec). With this fast decay, a simple feed-forward model yields reliable predictions of performance for various sequences of verniers and anti-verniers with individual vernier presentation times ranging from 5 to 60 msec (Figure 2, and 30 more unpublished conditions). Hence,  $\tau$  does not depend on the stimulus duration in this temporal range.

An important question concerns how the acquired information is read out from the fleeting visual retinotopic buffer and transferred to the more durable non-

retinotopic memory. One possibility is that integration starts with stimulus onset and the readout is flexibly coupled to the termination of the last stimulus rather than to its onset or a fixed readout time (Figure 6). The neural transients accompanying the temporal stimulus onset and termination can demarcate the beginning and the end of a stimulus. This is in accordance with physiological and theoretical considerations of visual persistence and masking (Macknik & Martinez-Conde, 2004; Macknik et al., 2000; Macknik & Livingstone, 1998; Francis, Grossberg, & Mingolla, 1994; Öğmen, 1993; Breitmeyer, 1984). If the timing of the stimuli is such that their transients cannot demarcate them into two different events, they fuse. Indeed, when vernier and anti-vernier are presented with an ISI longer than  $\sim 60$  msec, they no longer fuse but are perceived individually (see also Ma, Hamker, & Koch, 2005). In that case, the ISI allows transients that are sufficiently strong and temporally separate to segregate the individual stimuli.



**Figure 6.** Neural mechanism linking the fast retinotopic buffer and the nonretinotopic memory. In the first frame, a vernier and an anti-vernier are presented subsequently at one spatial position. If the timing of these stimuli is such that their transients do not demarcate them into two different events, they fuse. The fusion of vernier and anti-vernier in the retinotopic buffer can be described by an RC model based on decay of activity related to a spatial vernier offset. Because the vernier is presented first, its spatial offset information has decayed further than that of the anti-vernier. Therefore, the anti-vernier dominates. Its stimulus termination initiates the read-out and the fused feature information of the first frame is relayed to neural circuits that function as nonretinotopic visual memory. The nonretinotopic memory is postulated to decay at a much slower rate. The presentation of the second frame leads to a motion-grouping operation matching the three elements of the first frame to the corresponding three elements in the second frame. We suggest that this motion-grouping information determines how the information in the nonretinotopic memory becomes integrated with the information in the second frame. V = vernier; AV = anti-vernier;  $I_{a,b}$  = input to feature detector.



In the Ternus–Pikler display (Figure 4), the blank screen after the first frame produces a transient response initiating the readout of the fused feature information into a slow nonretinotopic memory. Hence, performance barely changes when ISI is varied. Likewise, feature information of the second frame is read out after its termination. Because the two frames are perceived as belonging to the same moving object, a correspondence of its elements is established according to motion grouping. Therefore, the stored information from the first frame is retrieved and integrated with the corresponding feature information of the second frame. Hence, the transients elicited by moving objects may trigger the transfer of the information from the fast retinotopic memory to the slow nonretinotopic memory, where prolonged motion integration computations can be carried out, thus avoiding smear.

Our combination of the retinotopic feature fusion paradigm with a nonretinotopic integration paradigm provides clear evidence for complex, dynamic processes that cannot be observed in single stimulus presentations. We show that there exists a fast stimulus processing buffer and that information of attended elements can be stored and integrated with elements at different locations and times. Interestingly, the integrated spatial offset is perceived as *one* offset without access to the spatial offsets of the individual elements that were integrated and that may have vanished. Accordingly, after counting the strokes of the clock, you may know what hour has come but hardly recall the single strokes. Several individual extrinsic events are condensed into one single intrinsic representation.

## APPENDIX: MODEL SPECIFICATION

The activation of neuron C can vary across a large range, whereas the psychophysical performance ranges between 0% and 100%. To map the activation of C to performance, we used a sigmoidal function as a linking hypothesis (equation 1), where C is the activation of neuron C at readout and  $\sigma$  is the steepness of the sigmoid function.

$$P(C) = \frac{100}{1 + \exp\left(-\frac{2C}{\sigma}\right)} \quad (1)$$

The sigmoid function maps activation around the middle range linearly to vernier. The sigmoid function applies a strong compression for activations far from the middle range as dominance asymptotes to 0 or 100.

The model has three free parameters: the decay rate constant  $\tau$ , the readout time  $r$ , and the slope of the linking sigmoid function  $\sigma$ . The data fit for all three parameters was obtained by means of a simplex search using the data of Experiment 2 and another 30 combinations of V-AV, V-AV-V, and V-AV-V-AV sequences with

individual vernier presentation times ranging from 5 to 60 msec (unpublished data). The resulting decay rate constant  $\tau$  was  $0.0291 \text{ msec}^{-1}$ , the readout time  $r$  was 70.8 msec after the termination of the last stimulus, and the slope  $\sigma$  of the linking sigmoid function was 101.5. For the entire data set of 34 combinations the predictions were quite accurate ( $r^2 = .89$ ).

In our simulations of the data of the second experiment, we set the input strength for the vernier and antivernier in the first frame to 1. The input strength for the vernier in the second frame was chosen such that the overall vernier dominance for the two frames together was close to 65%, as it was done in the experiment. For the model with continuous decay, this resulted in an input strength of 0.1. For the model in which the decay was frozen, the input strength for the vernier in the second frame was set to 0.15.

## Acknowledgments

This work was supported by the Swiss National Science Foundation (SNSF). We thank Greg Francis, Steve Macknik, and Günter Meinhardt for their comments and suggestions, and Marc Repnow for excellent technical help and useful discussions.

Reprint requests should be sent to Frank Scharnowski, EPFL SV BMI LPSY Station 15, CH-1015 Lausanne, Switzerland, or via e-mail: Frank.Scharnowski@epfl.ch.

## REFERENCES

- Bachmann, T. (2000). Microgenetic approach to the conscious mind. *Advances in consciousness research*, 25. Amsterdam: Johns Benjamins.
- Bergson, H. L. (1889). Essai sur les données immédiates de la conscience. In *La Bibliothèque de philosophie contemporaine* (pp. viii–185). Paris: Presses Universitaires de France.
- Bex, P. J., Edgar, G. K., & Smith, A. T. (1995). Sharpening of drifting, blurred images. *Vision Research*, 35, 2539–2546.
- Breitmeyer, B. G. (1984). *Visual masking: An integrative approach*. Oxford: Clarendon Press.
- Bridgeman, B., & Mayer, M. (1983). Failure to integrate visual information from successive fixations. *Bulletin of the Psychonomic Society*, 21, 285–286.
- Brincat, S. L., & Connor, C. E. (2006). Dynamic shape synthesis in posterior inferotemporal cortex. *Neuron*, 49, 17–24.
- Bullier, J., & Nowak, L. G. (1995). Parallel versus serial processing—New vistas on the distributed organization of the visual-system. *Current Opinion in Neurobiology*, 5, 497–503.
- Burr, D. (1980). Motion smear. *Nature*, 284, 164–165.
- Castet, E. (1994). Effect of the ISI on the visible persistence of a stimulus in apparent motion. *Vision Research*, 34, 2103–2114.
- Cattell, J. M. (1885). Über die Zeit der Erkennung und Benennung von Schriftzeichen, Bildern und Farben. *Philosophische Studien*, 2, 635–650.
- Chen, S., Bedell, H. E., & Ögmen, H. (1995). A target in real motion appears blurred in the absence of other proximal moving targets. *Vision Research*, 35, 2315–2328.
- Coltheart, M. (1980). Iconic memory and visible persistence. *Perception & Psychophysics*, 27, 183–228.

- Davidson, M. L., Fox, M. J., & Dick, A. O. (1973). Effect of eye-movements on backward masking and perceived location. *Perception & Psychophysics*, *14*, 110–116.
- Di Lollo, V. (1977). Temporal characteristics of iconic memory. *Nature*, *267*, 241–243.
- Dixon, N. F., & Hammond, E. J. (1972). The attenuation of visual persistence. *British Journal of Psychology*, *63*, 243–254.
- Efron, R. (1967). Duration of the present. *Annals of the New York Academy of Sciences*, *138*, 713–729.
- Efron, R. (1970a). Minimum duration of a perception. *Neuropsychologia*, *8*, 57–63.
- Efron, R. (1970b). Relationship between duration of a stimulus and duration of a perception. *Neuropsychologia*, *8*, 37–55.
- Efron, R. (1973). Conservation of temporal information by perceptual systems. *Perception & Psychophysics*, *14*, 518–530.
- Eriksen, C. W., & Collins, J. F. (1967). Some temporal characteristics of visual pattern perception. *Journal of Experimental Psychology*, *74*, 476–484.
- Farrell, J. E., Pavel, M., & Sperling, G. (1990). The visible persistence of stimuli in stroboscopic motion. *Vision Research*, *30*, 921–936.
- Francis, G. (1996). Cortical dynamics of visual persistence and temporal integration. *Perception & Psychophysics*, *58*, 1203–1212.
- Francis, G., Grossberg, S., & Mingolla, E. (1994). Cortical dynamics of feature binding and reset: Control of visual persistence. *Vision Research*, *34*, 1089–1104.
- Fröhlich, F. W. (1929). Die Empfindungszeit. In *Ein Beitrag zur Lehre von der Zeit- Raum- und Bewegungsempfindung* (pp. 1–366). Jena, Germany: Fischer Verlag.
- Haber, R. N. (1983). The impending demise of the icon—A critique of the concept of iconic storage in visual information-processing. *Behavioral and Brain Sciences*, *6*, 1–11.
- Hammett, S. T., Georgeson, M. A., & Gorea, A. (1997). Motion sharpening and blur. *Investigative Ophthalmology & Visual Science*, *38*, 1777–1777.
- Herzog, M. H., Parish, L., Koch, C., & Fahle, M. (2003). Fusion of competing features is not serial. *Vision Research*, *43*, 1951–1960.
- Hogben, J. H., & Di Lollo, V. (1985). Suppression of visible persistence in apparent motion. *Perception & Psychophysics*, *38*, 450–460.
- Irwin, D. E., Brown, J. S., & Sun, J. S. (1988). Visual masking and visual integration across saccadic eye-movements. *Journal of Experimental Psychology: General*, *117*, 276–287.
- Lamme, V. A. F., Zipser, K., & Spekreijse, H. (2002). Masking interrupts figure-ground signals in V1. *Journal of Cognitive Neuroscience*, *14*, 1044–1053.
- Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: An MEG study. *Nature Neuroscience*, *5*, 910–916.
- Ma, W. J., Hamker, F., & Koch, C. (2005). Neural mechanisms underlying temporal aspects of conscious visual perception. In H. Ögmen & B. G. Breitmeyer (Eds.), *The first half second: The microgenesis and temporal dynamics of unconscious and conscious visual processes* (pp. 275–294). Cambridge: MIT Press.
- Macknik, S. L., & Livingstone, M. S. (1998). Neuronal correlates of visibility and invisibility in the primate visual system. *Nature Neuroscience*, *1*, 144–149.
- Macknik, S. L., & Martinez-Conde, S. (2004). The spatial and temporal effects of lateral inhibitory networks and their relevance to the visibility of spatiotemporal edges. *Neurocomputing*, *58–60*, 775–782.
- Macknik, S. L., Martinez-Conde, S., & Haglund, M. M. (2000). The role of spatiotemporal edges in visibility and visual masking. *Proceedings of the National Academy of Sciences, U.S.A.*, *97*, 7556–7560.
- McRae, K., Butler, B. E., & Popiel, S. J. (1987). Spatiotopic and retinotopic components of iconic memory. *Psychological Research*, *49*, 221–227.
- Melcher, D., & Morrone, M. C. (2003). Spatiotopic temporal integration of visual motion across saccadic eye movements. *Nature Neuroscience*, *6*, 877–881.
- Metzger, W. (1932). Versuch einer gemeinsamen Theorie der Phänomene Fröhlichs und Hazelhoffs und Kritik ihrer Verfahren zur Messung der Empfindungszeit. *Psychological Research*, *16*, 176–200.
- Morgan, M. J., & Benton, S. (1989). Motion-deblurring in human vision. *Nature*, *340*, 385–386.
- Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.
- Ögmen, H. (1993). A neural theory of retino-cortical dynamics. *Neural Networks*, *6*, 245–273.
- Ögmen, H., Otto, T. U., & Herzog, M. H. (2006). Perceptual grouping induces real-time remapping of retinotopy in human vision. *Vision Research*, *46*, 3234–3242.
- Pikler, J. (1917). *Sinnesphysiologische Untersuchungen*. Leipzig, Germany: Barth.
- Pollack, I. (1973). Interaction effects in successive visual-displays—Extension of Eriksen–Collins paradigm. *Perception & Psychophysics*, *13*, 367–373.
- Ramachandran, V. S., Rao, V. M., & Vidyasagar, T. R. (1974). Sharpness constancy during movement perception—Short note. *Perception*, *3*, 97–98.
- Rubin, E. (1930). Bemerkungen zur Erwiderung Fröhlichs. *Psychological Research*, *13*, 289–292.
- Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs*, *74*, 1–29.
- Sugase, Y., Yamane, S., Ueno, S., & Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. *Nature*, *400*, 869–873.
- Super, H., Spekreijse, H., & Lamme, V. A. F. (2001a). A neural correlate of working memory in the monkey primary visual cortex. *Science*, *293*, 120–124.
- Super, H., Spekreijse, H., & Lamme, V. A. F. (2001b). Two distinct modes of sensory processing observed in monkey primary visual cortex (V1). *Nature Neuroscience*, *4*, 304–310.
- Ternus, J. (1926). Experimentelle Untersuchungen über phänomenale Identität. *Psychological Research*, *7*, 81–136.
- Thorpe, S. J., Delorme, A., & van Rullen, R. (2001). Spike-based strategies for rapid processing. *Neural Networks*, *14*, 715–725.
- Thorpe, S. J., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*, 520–522.
- Tong, J., Patel, S. S., & Bedell, H. E. (2005). Asymmetry of perceived motion smear during head and eye movements: Evidence for a dichotomous neural categorization of retinal image motion. *Vision Research*, *45*, 1519–1524.
- van Rullen, R., & Thorpe, S. J. (2001a). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artificial objects. *Perception*, *30*, 655–668.
- van Rullen, R., & Thorpe, S. J. (2001b). The time course of visual processing: From early perception to decision-making. *Journal of Cognitive Neuroscience*, *13*, 454–461.
- Westerink, J. H. D. M., & Teunissen, K. (1995). Perceived sharpness in complex moving images. *Displays*, *16*, 89–97.
- Yund, E. W., Morgan, H., & Efron, R. (1983). The micropattern effect and visible persistence. *Perception & Psychophysics*, *34*, 209–213.